Open Access   Review Article

AN UPDATED REVIEW ON DESIGN OF EXPERIMENT (DOE) IN PHARMACEUTICALS

Gaurav Gujral*, Devesh Kapoor, Manish Jaimini

[1] Research Scholar, Department of Pharmacy, Maharishi Arvind University, Jaipur, Rajasthan, India

[2] Dr. Dayaram Patel Pharmacy College, Sardarbaug, Station Road, Bardoli, Dist – Surat, Gujarat, India, Pin-394601

[3] Department of Pharmacy, Maharishi Arvind University, Jaipur, Rajasthan, India

ABSTRACT

DoE is a structured, organized method for determining the relationships among factors affecting a process and its output. It has been suggested that DoE can offer returns that are four to eight times greater than the cost of running the experiments in a fraction of the time that it would take to run one-factor-at-a-time experiments. It is always important before beginning experimentation to determine the objective of an experiment, and this is no different with DoE. Identifying objectives helps focus a team on its specific aims (scientific understanding of the task/problem in hand) over a period of time. It also helps indicate what resources are and assists in managing expectations from a study's outcome. DoE studies in support of QbD are often a delicate balance between delivering defined, high-quality products and meeting predetermined time, labor, and financial constraints.

Keywords: Quality by design (QbD), Design of experiment (DoE), Process analytical technologies (PAT), One factor at a time (OFAT), Quality risk management (QRM), Critical process parameters (CPP), Analysis of variance (ANOVA)

INTRODUCTION

The beginning of the twentieth century, Sir Ronald Fisher introduced the concept of applying statistical analysis during the planning stages of research rather than at the end of experimentation. When statistical thinking is applied from the design phase, it enables to build quality into the product, by adopting Deming's profound knowledge approach, comprising system thinking, variation understanding, theory of knowledge, and psychology. The pharmaceutical industry was late in adopting these paradigms, compared to other sectors. It heavily focused on blockbuster drugs, while formulation development was mainly performed by One Factor At a Time (OFAT) studies, rather than implementing Quality by Design (QbD) and modern engineering-based manufacturing methodologies. Among various mathematical modeling approaches, Design of Experiments (DoE) is extensively used for the implementation of QbD in both research and industrial settings. In QbD, product and process understanding is the key enabler of assuring quality in the final product. Knowledge is achieved by establishing models correlating the inputs with the outputs of the process[1].

Efficiency: Get more in information from fewer experiments

Focusing: Collecting the information that is really needed

There are 4 interrelated steps in building a design

1. Define the objective
2. Define the variable that will be controlled during experiment and their level /ranges of variation.
3. Define the variable that will be measured during experiment-Response variable
4. Choose among the variable standard design-the one that is compatible with the objective.

Table 1: The Table Below shows the various Designs available

Types of design	Screening	Factor influence	Optimization	Field of Use	No of Design variables
Full Factorial design	X	X		Study the effect of a lower number of design variables independently from each other, including interaction terms. *The only design that allows for categorical variables with 3 or more levels*	2-9
Fractional Factorial Design	X	X		Depending on the number of variables choose the *study lower order effects* independently from each other, or create a screening design aimed at finding the most important main effects among many,	3-13
Palackett-Burman Design	X			Economical *alternative to Fractional factorial design, studies main effects only*. Complex interaction effect.	8-35
Central Composite Design			X	Find the optimal levels of design variables by adding a few more experiments to a full fractional design. All *design variables must be continuous*	2-6
Box-Behnken Design			X	An alternative to central composite designs, when the optimum response is not located at the extremes of the experimental region and when previous results from a factorial design are not available. All design variables must be continuous	3-6
D-Optimal Design	X	X	X	Some design variables have multilinear constraints, and design is not orthogonal. Must be analysed with Partial least Squares Regressuion	2-9
Axial (Mixture) Design	X			Contains mixture variables only, design region is simplex. Only linear (first order) effects can be bound	3-30
Simplex-Lattice (Mixture) Design	X	X	X	Contains mixture variables only, design region is simplex. Tuneable lattice degree(order)	3-6 (9 if linear only)
SimplexCentroid (Mixture) Design			X	Contains mixture variables only, design region is simplex	3-6

Variables:

1. **Designed variables:** Variables with controlled variations are called design variables or factors. Design variable is completely defined by its name, its type: continuous category, its constraints: mixture, linear, its level.
2. **Response variables:** This is a type of Non-Designed variable, They are the measured outcome
3. **Non controllable variable:**
This second type of non designed variables refers to variables that are monitored.
May have an influence on the response variables and cannot be controlled or reliably fixed to a value e.g. Air humidity or Temperature

Continuous Vs Category variables:

Continuous variables:

They have a numerical value and can be measures quantitatively. E.g. Temperature, Concentration of ingredients. In this, variations are usually set within predefined range which goes from the lower to the higher. More levels between the extremes may be specified if the values are to be studied more specifically, If only 2 levels are specified-other necessary levels will be computed automatically.

Category Variables:

All the non continuous variables are category variables. Their level can be named, but not measured quantitatively. Bianry variables are special types of category variables that have only 2 levels-dichotomous.

Mixture variables:

When performing experiments where some ingredients are mixed according to the recipe, one may be in a situation where the amounts of the various ingredients cannot be varied independently from each other. In such case, one will need to use a special kind of design called a mixture design, and the design variables are called

mixture variables (or mixture components).An example of a mixture situation is blending concrete from the following three ingredients: cement, sand and water. If the percentage of water in the blend is increased by 10%, the proportions of one of the other ingredients (or both) will have to be reduced so that the blend still amounts to 100%.

However there are many situations where ingredients are blended, which do not require a mixture design. For instance in a water solution of four ingredients whose proportions do not exceed a few percentage, one may vary the four ingredients independently from each other and just add water at the end as filler. Therefore it is important to carefully consider the experimental situation before deciding whether the recipe being followed requires a mixture design or not.

Process Variables:

In a mixture situation, one may also want to investigate the effects of variations in some other design variables which are not themselves a components of the mixture. Such variables are called process variables. E.g. Temperature, Stirring rate, type of solvent, amount of catalyst.

Full-Factorial Design[2]:

Factorial experiments with two-level factors are used widely because they are easy to design, efficient to run, straightforward to analyze, and full of information. A full factorial design contains all possible combinations of a set of factors. This is the most fool proof design approach, but it is also the most costly in experimental resources. The full factorial designer supports both continuous factors and categorical factors with up to nine levels. Factorial designs with only two-level factors have a sample size that is a power of two (specifically 2 where f is the number of factors). When there are three factors have a sample size that is a power of three. N = L^k

Where   k = number of variables,
L = number of variable levels,
N = number of experimental trials,

Fractional Factorial Design:

Specific cases- with 2 level variables (continuous with upper and lower levels, and /or binary variables)-one can define fraction of full factorial design. It enables the investigation of as many design variables are chosen full factorial design with fewer experiments. These designs might be set up by selecting half the experimental runs of the original design.

Example of Fractional factorial design

Four design variables-A, B, C, D. Lower and upper levels are coded '-'and '+'respectively.

First the full factorial design is built with only 3 variables A, B & C (2^3) as shown below.

Table 2: Full factorial design is built with only 3 variables

Experiment	A	B	C
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

Full Factorial Design 2^3

Fractional Factorial Design: In the table below, additional columns are generated, which are computed from the products of the original 3 columns A, B, C. these additional columns represent the interactions between the design variables.

Table 3: Description of full factorial design 2^3

Experiment	A	B	C	AB	AC	BC	ABC
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

Full factorial design 2^3 with interaction column

Confounding:

Confounding is the side effect of the Fractional factorial. Confounding means that some effects cannot be studied independently of each other.

Resolution of a full factorial design:

How well a fractional –factorial design avoids confounding is expressed through its resolution. The 3 most common cases are as follows.

1. Resolution III Design : Main effects are confounded with two-factor interaction
2. Resolution IV Design: Main effects are free of confounding with two-factor interaction, but two factor interactions are confounded with each other.

3. Resolution V Design: Main effects and two-factor interactions are free of confounding with each other; however some two factor interactions are confounded with three factor interactions.
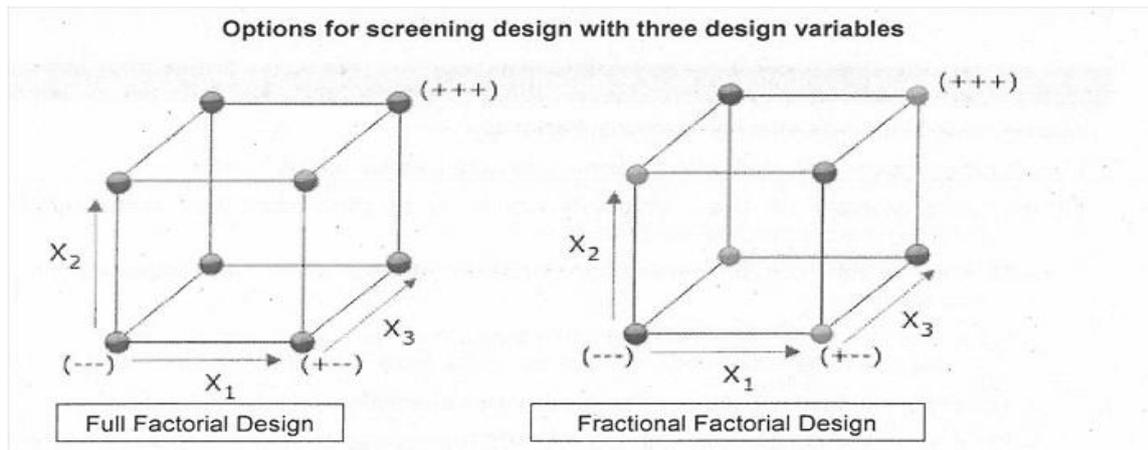

Figure 1: Description of full factorial design and fractional factorial design

Plackett-Burman (PB) design:

Plackett-Burman (PB) designs are used for screening experiments because, in a PB design, main effects are, in general, heavily confounded with two-factor interactions. The PB design in 12 runs, for example, may be used for an experiment containing up to 11 factors.

In this experimental objective should be the study of main effects only. It is very economical as they require only 1to4 more experiments than the number of design variables. The interaction between the factors is considered negligible.

Run	A	B	C	D	E	F	G	H	J	K	L
1	+	−	+	−	−	−	+	+	+	−	+
2	+	+	−	+	−	−	−	+	+	+	−
3	−	+	+	−	+	−	−	−	+	+	+
4	+	−	+	+	−	+	−	−	−	+	+
5	+	+	−	+	+	−	+	−	−	−	+
6	+	+	+	−	+	+	−	+	−	−	−
7	−	+	+	+	−	+	+	−	+	−	−
8	−	−	+	+	+	−	+	+	−	+	−
9	−	−	−	+	+	+	−	+	+	−	+
10	+	−	−	−	+	+	+	−	+	+	−
11	−	+	−	−	−	+	+	+	−	+	+
12	−	−	−	−	−	−	−	−	−	−	−

Figure 2: Plackett-Burman design for 12 runs and up to 11 two-level factors

Response Surface Design[4]:

Response surface design is a set of advanced design of experiments (DoE) techniques that help you better understand and optimize your response. Response surface design methodology is often used to refine models after you have determined important factors using screening designs or factorial designs; especially if you suspect curvature in the response surface.
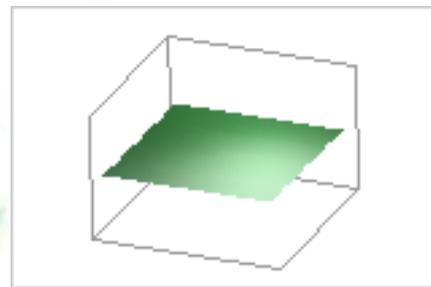

Figure 3: Response surface with no curvature

Figure 4: Response surface with curvature

The difference between a response surface equation and the equation for a factorial design is the addition of the squared (or quadratic) terms that lets you model curvature in the response, making them useful for:

- Understanding or mapping a region of a response surface. Response surface equations model how changes in variables affect a response of interest.
- Finding the levels of variables that optimize a response.
- Selecting the operating conditions to meet specifications.

There are two main types of response surface designs:
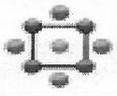
Central Composite designs (CCD)

Central Composite designs can fit a full quadratic model. They are often used when the design plan calls for sequential experimentation because these designs can include information from a correctly planned factorial experiment. A central composite design is the

most commonly used response surface designed experiment. Central composite designs are a factorial or fractional factorial design with center points, augmented with a group of axial points (also called star points) that let you estimate curvature. You can use a central composite design to:

- Efficiently estimate first- and second-order terms.
- Model a response variable with curvature by adding center and axial points to a previously-done factorial design. Central composite designs are especially useful in sequential experiments because you can often build on previous factorial experiments by adding axial and center points.

Properties of Central Composite designs:

Rotability: We do not know the position of the response surface optimum; we try to ensure that the prediction error is same for any point at the same distance from the centre of the design. This property is called rotability.

Circumscribed Central Composite Design (CCD)	
Faced Central Composite Design (CCF)	
Inscribed Central Composite Design (CCI)	
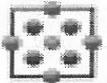
**Figure 5: Different types of CCD**

Efficiency of CCD:

Depending on the constraints of the experiments and the accuracy to achieve, select the appropriate CC design using the following table.

Table 4: Different designs with different number of levels with accuracy of estimates

Design	Number of levels	Uses point outside high and low variables	Accuracy of estimates
Circumscribed	5	Yes	Good over entire design space
Inscribed	5	No	Good over central subset of the design space
Faced	3	No	Fair over entire design space, poor for pur quadratic coefficients
Box Behnken	3	No	Good over entire design space, more uncertainly on the edge of the design area

Box-Behnken designs:

A Box-Behnken design is a type of response surface design that does not contain an embedded factorial or fractional factorial design. Box-Behnken designs usually have fewer design points than central composite designs, thus, they are less expensive to run with the same number of factors. They can efficiently estimate the first- and second-order coefficients; however, they can't include runs from a factorial experiment. Box-Behnken designs always have 3 levels per factor, unlike central composite designs which can have up to 5. Also unlike central composite designs, Box-Behnken designs never include runs where all factors are at their extreme setting, such as all of the low settings. Box-Behnken designs have treatment combinations that are at the midpoints of the edges of the experimental space and require at least three continuous factors. The following figure shows a three-factor Box-Behnken design. Points on the diagram represent the experimental runs that are done:
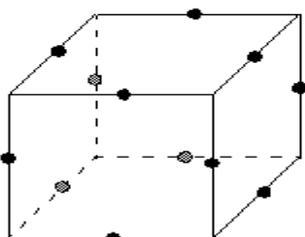
Figure 6: Three-factor Box-Behnken design

These designs allow efficient estimation of the first- and second-order coefficients. Because Box-Behnken designs often have fewer design points, they can be less expensive to do than central composite designs with the same number of factors. However, because they do not have an embedded factorial design, they are not suited for sequential experiments. Box-Behnken designs can also prove useful if you know the safe operating zone for your process. Central composite designs usually have axial points outside the "cube." These points may not be in the region of interest, or may be impossible to conduct because they are beyond safe operating limits. Box-Behnken designs do not have axial points, thus, you can be sure that all design points fall within your safe operating zone. Box-Behnken designs also ensure that all factors are not set at their high levels at the same time.

Choice of experimental design[5]:

The most important part of a DoE process, choosing an appropriate experimental design, is critical for the success of the study. The choice of experimental design depends on a number of aspects , including the nature of the problem and/or study (e.g., a screening, optimization, or robustness study), the factors and interactions to be studied (e.g., four, six, or nine factors, and main effects or two-way interactions), and available resources (e.g., time, labour, cost, and materials). Using previous knowledge of a product or previous experiments to identify possible interactions among the input process parameters before performing an

experiment also plays a key part in selecting an appropriate experimental design.

Statistical Analysis (Model Selection, Residual Analysis, and Transformation of Response) [5]:

Once data have been collected according to the chosen design, the results should be analysed using statistical methods so that objective conclusions can be drawn. Many software packages are available to assist, including those that help users choose a design to those that perform statistical analysis, report results, and generate a mathematical model. One such model is the ANOVA, which is a statistical method based on the F-test to assess the significance of model terms. Once the appropriateness of those terms and the overall model satisfies an ANOVA check, the next step is to determine what cannot be modelled. This is done by residual analysis.

Software and Statistical Awareness[5]:

Good DoE software helps users follow the regressive modelling approach. It should guide them in carefully choosing model terms on the basis of graphical tools and statistics, and it should verify a model and its significance based on statistics in addition to verifying unaccounted residuals. Graphical tools play a key part in understanding and presenting statistical analysis results, so make sure that they deliver a smart way to diagnose, analyse, predict, and present the results in two and three dimensions.

.

A systematic application of DoE facilitates the identification of CPPs and their relationship to CQAs, leading to the development of a design space. In combination with quality risk management (QRM) and process analytical technologies (PAT), these help companies maintain good manufacturing control and consistency, ultimately guaranteeing the quality of their drug products.

CONCLUSION

Nowadays, much of the scientific basis is already in place for the implementation of QbD. So, the Statistical optimization for pharmaceutical scientist is to define the formulation with optimum characteristics. Statistical optimization can also provide solutions to larger-scale manufacturing problems, which occasionally arise. Importantly, statistical optimization experimentation and analysis provides strong assurances to Regulatory Agencies regarding superior product quality.

REFERENCES

1. Design of experiments (DoE) in pharmaceutical development, Drug Dev Ind Pharm. 2017 Jun; 43(6):889-901.
2. Chowdary K.P.R., Ravi Shankar K., Optimization of pharmaceutical product formula-tion by factorial designs: case studies, Journal of pharmaceutical research, 2016; 15(4): 105-109
3. The Design of Optimal Multifactorial Experiments" in *Biometrika* (vol. 33).
4. Support.minitab.com
5. Shivare M, Mc Creath G, Practical consideration for DOE Implementation in Quality by design, Bioprocess, Technical, 2010.